RSAC 2025 Conference

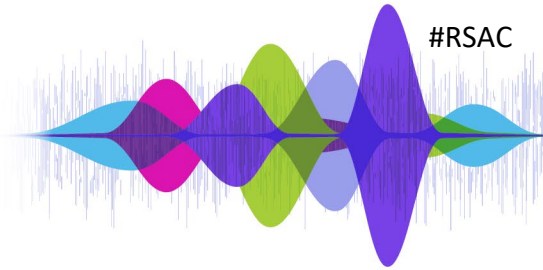
Many Voices.
One Community.

SESSION ID: FRP-W08

The Dark Side of Innovation: Generative AI in Cybercrime

John Fokker

Head of Threat Intelligence
Trellix



Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

Two decades after Edison switched on the light bulb, only 3% of U.S. businesses used electricity”

Ajay Agrawal, Joshua Gans, and
Avi Goldfarb in Harvard Business Review

Trellix

Gen-AI usage by Cybercriminals

What is happening now...

Project “ChatGPT in Jabber”

ChatGPT in Jabber

xatab · Сегодня в 00:49

xatab
NTD$
Premium
Регистрация: 02.01.2024
Сообщения: 79
Реакции: 35
Депозит: 1 ₿

Сегодня в 00:49

1. Добавляем в контакты chatgpt@thesecure.biz
2. Пишем слово "hunter"
3. Добавляем API key сообщением set_key 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
4. Наслаждаемся общением с ChatGPT 4.0 через Jabber

Кому полезно, дайте обратную связь! 😎

Жалоба

reqwest

Python

```
1. Add to your contacts chatgpt@thesecure.biz
2. Write a keyword "hunter"
3. Add API key in the message: set_key <OPENAI_API_KEY>
4. Enjoy the conversation with  ChatGPT 4.0 via Jabber

If useful share your feedback!
```

AI Integrated Web-Scanner

ИИ в Пентесте: Автоматизированный Веб-Сканер с Интеграцией ИИ

hackeryaroslav · 19.01.2024

1 2 Вперёд ▸ Отслеживать

hackeryaroslav
(L3) cache
Пользователь
Регистрация: 11.09.2023
Сообщения: 293
Реакции: 182

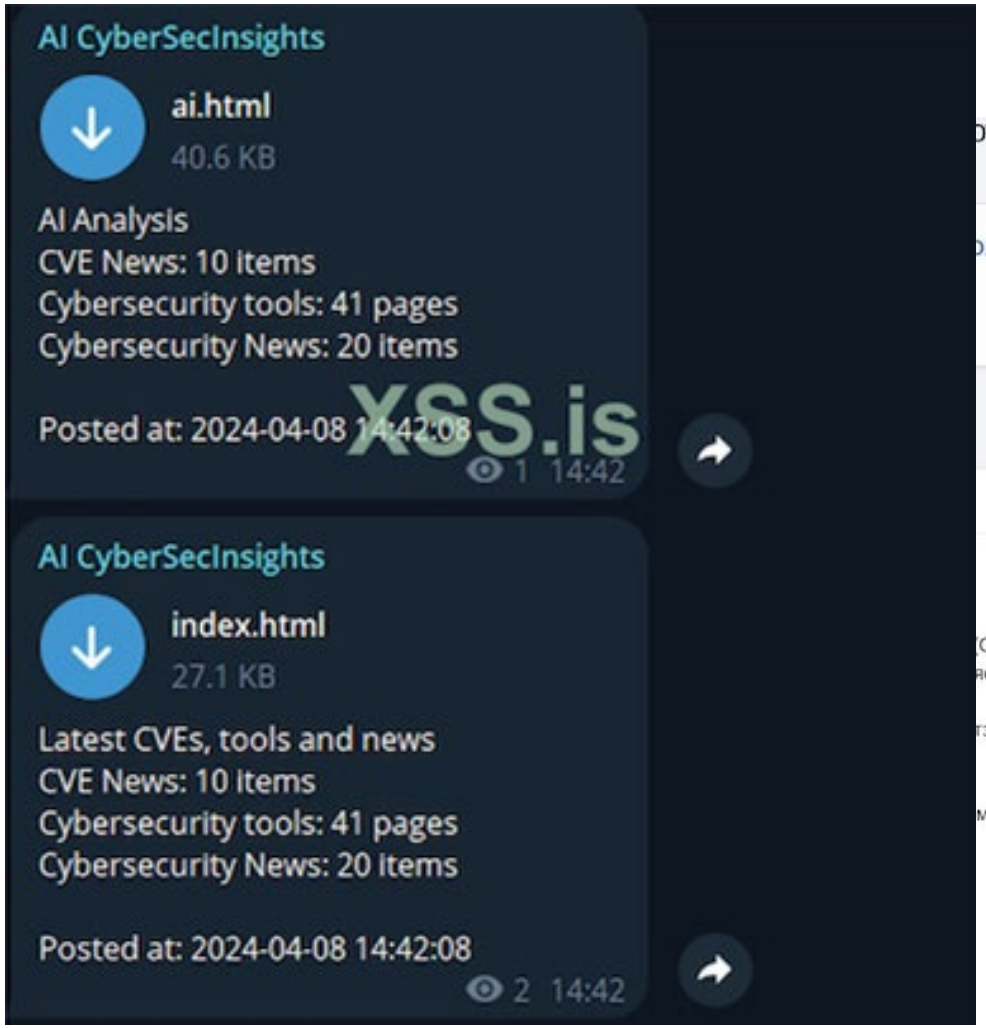
19.01.2024 #1

Авторство: hackeryaroslav

Источник: xss.is

Автоматизированный Веб-Сканер v1.0 (планирую добавлять новые идеи и доработки, буду оказывать активную поддержку, так что следите за обновлениями!). Сканер представляет собой средство обнаружения уязвимостей в веб-приложениях.

Gemini AI powered Vuln Intelligence



CVE-2024-3085 A vulnerability classified as critical has been found in PHPGurukul Emergency Ambulance Hiring Portal 1.0. Affected is an unknown function of the file /admin/login.php of the component Admin Login Page. The manipulation of the argument username leads to sql injection. It is possible to launch the attack remotely. The exploit has been disclosed to the public and may be used. VDB-258678 is the identifier assigned to this vulnerability.

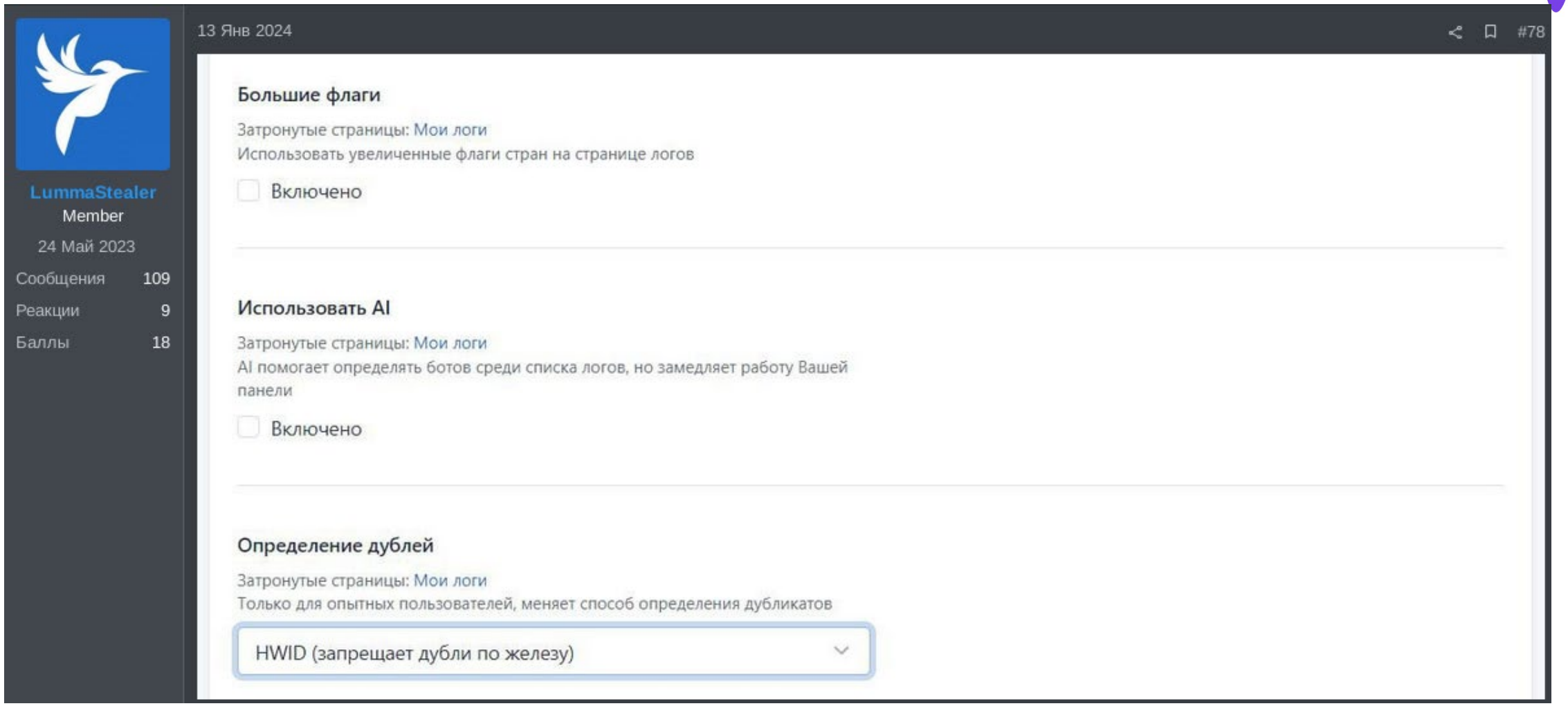
AI Feedback

Summary Feedback: **CVE Summary Analysis** **Vulnerability Overview:** * The vulnerability is classified as **critical**, indicating a high potential for severe impact. * It affects **PHPGurukul Emergency Ambulance Hiring Portal 1.0**. * The vulnerable component is the **Admin Login Page** (specifically, the /admin/login.php file). * The vulnerability allows for **SQL injection** attacks, which can manipulate database queries. **Technical Details:** * The vulnerability is triggered by manipulating the **username** argument of an unknown function in the /admin/login.php file. * By crafting malicious username input, an attacker can inject arbitrary SQL commands into the database. **Impact:** * SQL injection attacks can allow attackers to: * Access and modify sensitive data * Execute arbitrary commands on the database * Perform data exfiltration **Exploitation:** * The vulnerability can be exploited **remotely**, indicating that an attacker does not need direct access to the affected system. * The exploit has been **disclosed to the public**, meaning it is widely known and may be actively exploited. **Identifier:** * The vulnerability has been assigned the identifier **VDB-258678**. **Recommendations:** * Update PHPGurukul Emergency Ambulance Hiring Portal to the latest version to address the vulnerability. * Implement input validation and sanitization to prevent SQL injection attacks. * Regularly monitor for security updates and apply them promptly.

Detailed Feedback: **CVE ID:**CVE-2024-3085 **Description:** A critical vulnerability exists in PHPGurukul Emergency Ambulance Hiring Portal 1.0. A vulnerability classified as critical has been found in PHPGurukul Emergency Ambulance Hiring Portal 1.0. Affected is an unknown function of the file /admin/login.php of the component Admin Login Page. The manipulation of the argument username leads to sql injection. It is possible to launch the attack remotely. The exploit has been disclosed to the public and may be used. VDB-258678 is the identifier assigned to this vulnerability. **References:** * [MISC:Submit #306958 | PHPGurukul Emergency Ambulance Hiring Portal 1.0 SQL Injection](https://vuldb.com/?submit.306958) * [URL:https://vuldb.com/?submit.306958](https://vuldb.com/?submit.306958) * [MISC:VDB-258678 | CTI Indicators (IOB, IOC, TTP, IOA)](https://vuldb.com/?ctiid.258678) * [URL:https://vuldb.com/?ctiid.258678](https://vuldb.com/?ctiid.258678) * [MISC:VDB-258678 | PHPGurukul Emergency Ambulance Hiring Portal Admin Login Page login.php sql injection](https://vuldb.com/?id.258678) * [URL:https://vuldb.com/?id.258678](https://vuldb.com/?id.258678) * [MISC:https://github.com/dhabaleshwar/Open-Source-Vulnerabilities/blob/main/eahp_sqli.md](https://github.com/dhabaleshwar/Open-Source-Vulnerabilities/blob/main/eahp_sqli.md) * [URL:https://github.com/dhabaleshwar/Open-Source-Vulnerabilities/blob/main/eahp_sqli.md](https://github.com/dhabaleshwar/Open-Source-Vulnerabilities/blob/main/eahp_sqli.md) **Date Record Created:** 20240329 **Phase (Legacy):**Assigned (20240329) **Vulnerability Type:** SQL Injection **Affected Software Versions:**PHPGurukul Emergency Ambulance Hiring Portal 1.0 **Additional Information:** **CNA (Common Names):** **VulDB** **Discovery Date:** Not provided **Disclosure Date:** Not provided **Exploit Availability:** Public **Impact:** Critical **Attack Vector:** Remote **Attack Complexity:** Low **Authentication:** Not required **Severity Rating:** High **Remediation:** Not provided

XSS.is

Gen-AI adoption in Infostealers



AI-Powered Credential Validation

- AI is used to clean, deduplicate and verify Stolen credentials.

AI Voice Cloning Services

[ВПЕРВЫЕ НА РЫНКЕ] ИИ звонки | WhatsApp & Телефон | Любой голос, любой язык

PatrickCompany · 28.01.2025 · звонки по миру · прозвон · прозвонщик

В ЭТОЙ ТЕМЕ МОЖНО ИСПОЛЬЗОВАТЬ АВТОМАТИЧЕСКИЙ ГАРАНТ!

Новая сделка

Отслеживать

PatrickCompany
CD-диск
Пользователь
Регистрация: 28.08.2023
Сообщения: 13
Реакции: 2

28.01.2025 #1

Цена: $$$
Контакты: https://href.li/?https://t.me/PatrickCompanySales

Представляем Вам инновационное решение в мире телефонных звонков на базе искусственного интеллекта

Наш продукт создан для автоматизации исходящих и входящих звонков , интегрируемых с любыми мессенджерами и IP-телефонией. Это мощный инструмент, который поможет вашему бизнесу оптимизировать процессы коммуникации и повысить эффективность работы.

Наша технология базируется на основе ИИ и может адаптироваться под нестандартные вопросы и ситуации, также, при необходимости, она может выходить за рамки контекста, обеспечивая гибкость и точность в решении различных задач.

Кроме того, голоса ИИ созданы на основе реальных голосов людей и практически не отличимы от натуральной речи. По нашему опыту, собеседники никогда не понимали что они разговаривают с ИИ.

Ransomware actors using GPT

- BlackBasta used ChatGPT during their attack to write a fake letter to calm a user that noticed their activity.

- BlackBasta Coders used ChatGPT to debug their code for errors and convert C# code to Python.

```
2024-02-21T23:53:51 @usernamenn:matrix.bestflowers247.online    the guy noticed something suspicious
2024-02-21T23:53:55 @usernamenn:matrix.bestflowers247.online    when I connected to his computer
2024-02-21T23:54:00 @usernamenn:matrix.bestflowers247.online    started panicking opened chat
2024-02-21T23:54:02 @usernamenn:matrix.bestflowers247.online    with me
2024-02-21T23:54:15 @usernamenn:matrix.bestflowers247.online    I quickly tricked ChatGPT and asked it to write me a fake letter
2024-02-21T23:54:19 @usernamenn:matrix.bestflowers247.online    believable one
2024-02-21T23:54:21 @usernamenn:matrix.bestflowers247.online    Upon encountering issues with the domain controller, a thorough examination of the hosts was conducted by a specialized technical professional.
The investigation encompassed an in-depth scrutiny of the network infrastructure, focusing on the integrity and functionality of each host within the domain. This encompassed a meticulous review of network
configurations, system logs, and performance metrics to identify any anomalies or irregularities that could potentially impede the domain controller's operations. The examination was executed with meticulous
attention to detail, employing diagnostic tools and methodologies tailored to the specific nuances of domain controller operations. Through this comprehensive analysis, the technical specialist sought to
pinpoint and rectify any underlying issues affecting the stability and performance of the domain controller, thereby ensuring the seamless operation of the network environment.
2024-02-21T23:54:28 @usernamenn:matrix.bestflowers247.online    I sent it to him and he calmed down
...
2024-02-22T00:11:45 @usernamenn:matrix.bestflowers247.online    he said no problems thanks for explaining))
2024-02-22T00:11:55 @usernamenn:matrix.bestflowers247.online    damn what a persistent guy we got
...
2024-02-22T00:12:59 @usernamenn:matrix.bestflowers247.online    yeah fuck I went to his PC, and somehow got thrown into his session, he saw that his computer was being controlled and started all these
```

State Sponsored Usage of GenAI

Trellix

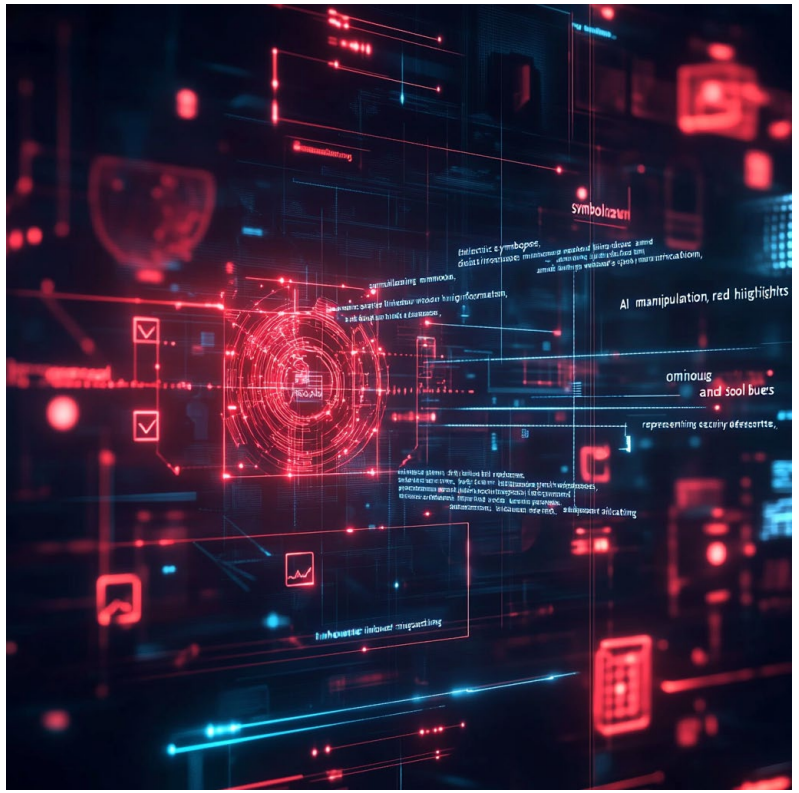
Earlier in 2024…

Groups linked to:

- Russia
- North Korea
- China
- Iran

OpenAI Threat Intelligence Reports (Feb 2025)

- Gen AI usage to create content for (mis)information operations by Nation states.
- First time AI generated articles apear in mainstream media in a different Geo.
- Pig-butchering or Romance Scam content creation.
- Iranian Linked Information Operations



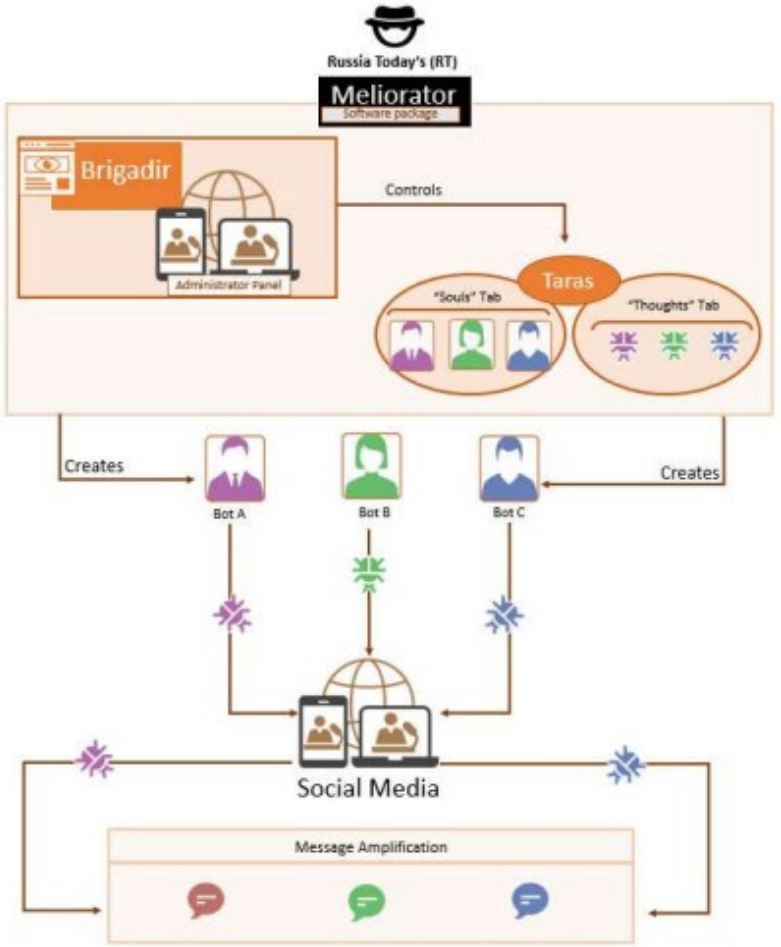
https://cdn.openai.com/threat-intelligence-reports/disrupting-malicious-uses-of-our-models-february-2025-update.pdf

State-sponsored use of AI

Meliorator: Overview

AI-Powered Bot Farm

- Generates and manages realistic social media personas called "souls."
- Components:
  - **Brigadir:** Administrator panel for managing the bot farm.
  - **Taras:** Back-end seeding tool for distributing content.
- Avoids Detection
- Multiple Platforms.
- Disinformation campaigns, notably around the Russia-Ukraine conflict.



State Sponsored Usage of GenAI

State Sponsored Job infiltration through Deep Fakes

Operation Dream Job 2.0

- A possible DPRK operative attempted to infiltrate KnowBe4 posing as an IT AI worker.
- The operative used a stolen US-based identity and AI-enhanced the photo.

Modus Operandi:

- Operatives use "IT mule laptop farms" and VPNs to work remotely.
- Aim to earn money and fund North Korean programs.

Preventive Measures:

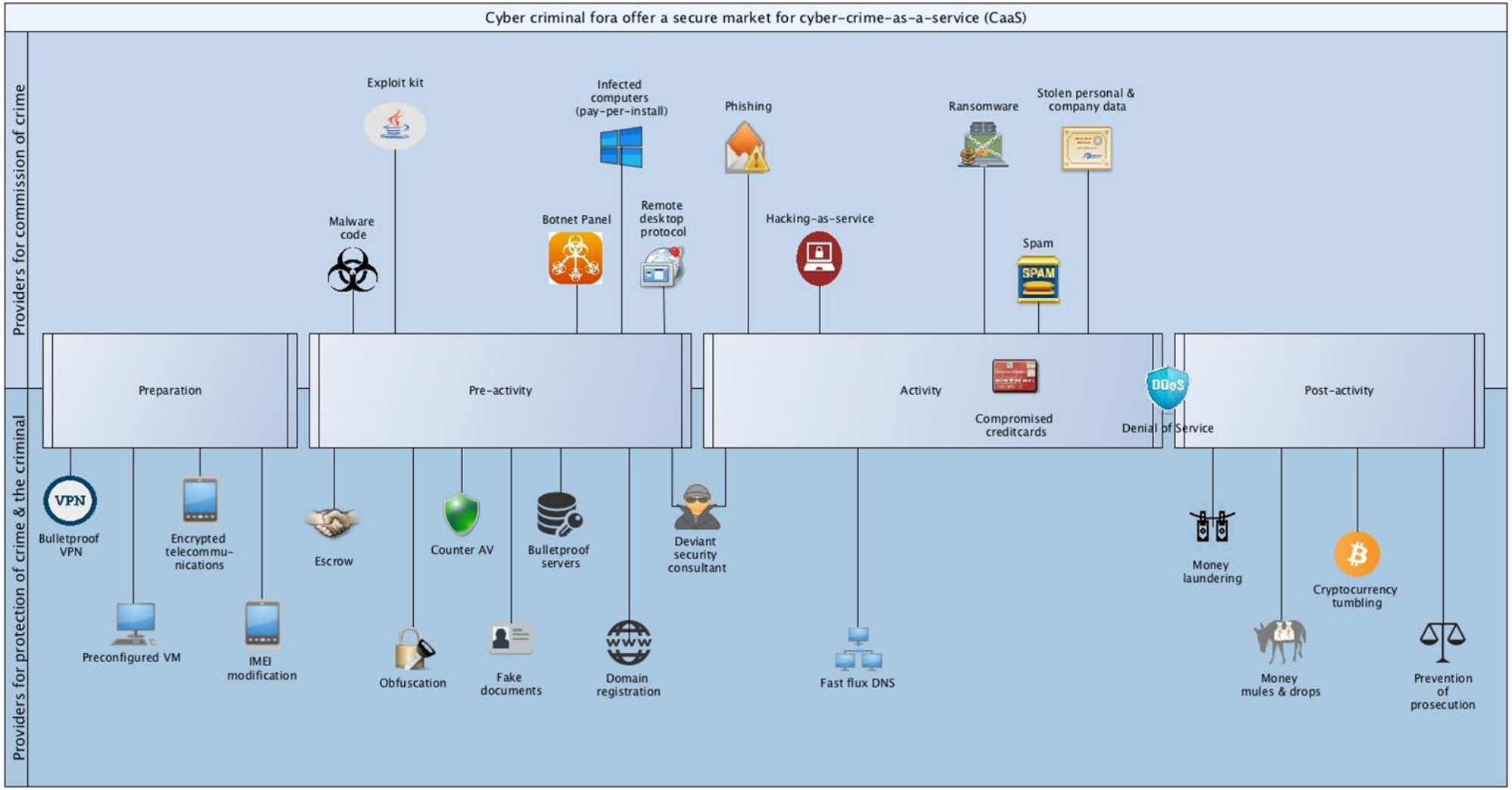
- Enhanced background checks and continuous security monitoring.
- Verify physical location and scrutinize resume details.

Going from AI Innovation to AI Transformation

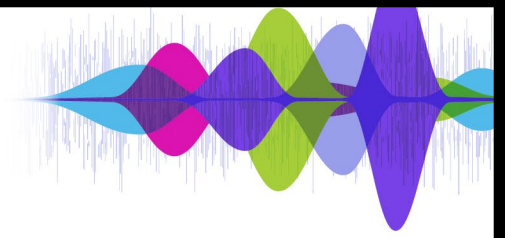
Two decades after Edison switched on the light bulb, only 3% of U.S. businesses used electricity”

Ajay Agrawal, Joshua Gans, and
Avi Goldfarb in Harvard Business Review

Cyber Criminal underground Network of services



An AI System Change takes time



Conclusions and Applying knowledge

Rise of the machines by the speed of human adoption

Conclusion and applying knowledge

Conclusion

- Cyber-criminals faced with similar challenges
- Going from Innovation to Transformation takes time (lesser now)
- Weak passwords/ unpatched systems still make a target Rich Environment.

Applying knowledge

- Stay Informed on AI Advancements
- Proactive Security
- Collaborative Defense
- Ethical AI Use

RSAC™ 2025 Conference

Many Voices.
One Community.